

## **The Reconstruction of Ancestral Character States**

### Authors:

Ted R. Schultz  
(Corresponding author)  
Department of Entomology  
Comstock Hall  
Cornell University  
Ithaca, NY 14853 U.S.A.  
607-255-4564 and 607-272-6209  
ts15@cornell.edu

Reginald B. Cocroft  
Department of Neurobiology and Behavior  
Seeley G. Mudd Hall  
Cornell University  
Ithaca, New York 14853 U.S.A.

Gary A. Churchill  
Biometrics Unit  
Department of Plant Breeding and Biometry  
Warren Hall  
Cornell University  
Ithaca, New York 14853 U.S.A.

Right running head: Ancestral state reconstruction

Left running head: T. R. Schultz et al.

Key words: Adaptation, character mapping, homoplasy, homology, phylogeny, parallelism

Uncovering the path of evolution through the distribution of features in extant species has been a primary tool of evolutionary biology since its inception (Darwin, 1859). This comparative, historical approach is especially important for the study of adaptive traits, because only within a phylogenetic framework do such pivotal concepts as convergence, parallelism, and adaptive radiation become meaningful. Early evolutionary studies of behavioral and ecological traits were explicitly phylogenetic (e.g., Whitman, 1899; Heinroth, 1911; Chapin, 1917; Lorenz, 1941; Lack, 1947; Tinbergen, 1951), and during the past decade evolutionary biology has seen a resurgence of interest in this approach (e.g., Ridley, 1983; Coddington, 1988; Carpenter, 1989; Brooks and McLennan, 1991; Baum and Larson, 1991; Harvey and Pagel, 1991). This increased appreciation for the role of historical processes can be attributed at least in part to recent major conceptual and methodological advances in the field of systematics (Hennig, 1966; Eldredge and Cracraft, 1980; Nelson and Platnick, 1981; Wiley, 1981; Felsenstein, 1982; Farris, 1983; Ax, 1987; Sober, 1988; Hillis et al., 1994).

Responding to this trend, Frumhoff and Reeve (1994) examine both the usefulness and reliability of phylogenetic information in the study of evolutionary adaptation. In addition to scrutinizing specific methodological proposals, these authors conclude that the inference of ancestral character states from the distribution of character states present in descendant species is highly error-prone. Because the inference of ancestral character states is central to the phylogenetic study of evolutionary patterns, this criticism, if accurate, has important negative

consequences for historical studies of behavioral, ecological, and morphological evolution.

Here we focus on the issue of whether or not the distribution of character states within extant species provides information about ancestral states. We begin by arguing that evolutionary homology provides the central concept for understanding this issue. We extend the model contributed by Frumhoff and Reeve (1994) for the neglected problem of error in ancestral character-state inference, focusing on the two evolutionary mechanisms identified by those authors: random and concerted homoplasy. By imposing confidence levels, we demonstrate that Frumhoff and Reeve's (1994) reported error rates are unduly pessimistic. We discuss the parameters that must be quantified in order to judge the likelihood of the general mechanism for parallel evolution proposed by those authors, and conclude with a discussion of how the model might be made more realistic.

### Homology

The difference between homology and homoplasy lies at the heart of the problem of reconstructing ancestral character states. In phylogeny reconstruction the assignment of the same character state to two or more species constitutes a hypothesis of homology, i.e., the character states are hypothesized to be similar due to an unbroken chain of inheritance from a common ancestor that also possessed that state. When a number of such presumed homologies are combined into a phylogenetic analysis, a cladogram is identified in which the number of evolutionary origins of hypothesized homologies is minimized or in which their joint probability is maximized (as the sole criterion in "maximum parsimony" methods, or under additional a priori constraints in various other methods).

When most a priori homology assessments are correct, distributional congruence among characters will be high. This is because, due simply to Darwinian "descent with modification" (Darwin, 1859: 123), we expect homologous characters to be correlated in accord with the same phylogenetic pattern. But even in cases where the proportion of mistaken homology assessments (homoplasies, i.e., unrecognized convergences, parallelisms, and reversals) is relatively high, we do not expect to be systematically misled by such characters as a general rule of evolution, because the distribution of homoplasious characters in extant species is predicted to be, on average, random with respect to one another and with respect to the distribution of homologous characters. In other words, we expect homologous characters to produce a phylogenetic "signal" against a background of homoplasious "noise" (Hillis, 1991; de Pinna, 1991; Patterson, 1982). (A possible exception arises under the parsimony criterion in the case of "long branch attraction" [Felsenstein, 1978; Hendy and Penny, 1989], which requires conditions that depart from the equal branch lengths modeled by Frumhoff and Reeve [1994], and that at any rate are unlikely to apply to the complex morphological, behavioral, and ecological characters discussed and modeled here.)

Contrary to these assumptions, Frumhoff and Reeve (1994) propose that homoplasious characters can be expected with high probability to be distributed in pseudophylogenetic patterns that are positively correlated with the distributions of genuine homologies. They identify two mechanisms responsible for generating this phenomenon: random homoplasy and concerted homoplasy (parallel evolution), which we discuss in turn below.

Pseudophylogenetic patterns produced by random homoplasy

In phylogeny reconstruction it is generally assumed that, when all members of a monophyletic group of species possess the same character state, the unobserved ancestor(s) imbedded within that clade also possessed that state. (The algorithm for inferring ancestral states under the parsimony criterion, known as character state optimization, is formalized in Farris [1970], Fitch [1971], Sankoff and Rousseau [1975], Sankoff and Cedergren [1983], and Swofford and Maddison [1987], among others.) For example, the consistent presence of a vertebral column in 39,500 extant vertebrate species leads us to hypothesize that their common ancestor, as well as all subsequent ancestors, also possessed this anatomical feature, i.e., that the vertebral column had a single evolutionary origin rather than up to 39,500 separate ones.

Challenging this logic, Frumhoff and Reeve (1994: 174) suggest that the inference of an ancestral character state can be "extremely error-prone" and that "even as the number of sister taxa sharing a common character state increases, there may remain a substantial probability that their common ancestor possessed an alternative character state." In particular, using a Markov model (with specific values reproduced in our Matrix 1; see Appendix) for a two-state character in which one state is found to be fixed for all the species in a polytomy (e.g., state A in Figure 1), they demonstrate that the expected error can be very high when 1 or more changes per branch are expected to have occurred. For instance, when the expected number of transitions within each descendant lineage is 4, error rates remain as high as nearly 0.5 even when there are as many as 20 descendants and all are fixed for the same character state. When the expected number of transitions is 1, error rates do not decrease to the 0.05 level until approximately 11 descendant species are present (Frumhoff and Reeve, 1994: 174, Figure 2). This argument makes intuitive

sense. Rapidly evolving characters retain little information on history, because only a single transition is required to obliterate homology.

The critical concept that must be added to Frumhoff and Reeve's (1994) analysis is that rapidly evolving characters governed by the Markov process specified in Matrix 1 are unlikely to generate a pattern of character-state fixation for a single state (rather than a mosaic of both states) in all  $N$  descendant species, particularly when  $N$  is large (Figure 2). When a pattern of fixation is encountered (e.g., character state A in Figure 1), it provides strong evidence that the character does not fit this model of rapid evolution. In fact, the probability of fixation for the observed state when any amount of accumulated change approaching 0.5 or more transitions per branch has occurred is significantly low when a clade contains more than 4 species. It is clear that, when a such a pattern is encountered, it is extremely unlikely that this amount of change has taken place. This additional information about the rate of evolution must be taken into account when calculating the probability of expected error in reconstructing the ancestral state (see Appendix).

It seems reasonable to consider only the values for the expected number of transitions per branch (positively correlated with "time," or  $t$ ; see Appendix) that are consistent with the observation of fixation in all  $N$  species. The approach used here for estimating the probability of error in ancestral character-state reconstruction is first to place an upper bound on the expected number of transitions per branch and, second, to calculate the error probability corresponding to that value. Since error probability increases as expected number of transitions increases, this approach provides a conservative upper bound within the context of the model's assumptions. For instance, if we choose to consider only those expected numbers of transitions per branch that will generate a pattern of fixation in 5 species (Figure 1)

with a probability of 0.95, this translates into an upper bound of 0.179 expected transitions per branch. The corresponding maximum error associated with this number of expected transitions is 0.081 (Figure 3.2).

By this reasoning, the range of evolutionary rates considered by Frumhoff and Reeve (1994) in their Figure 2 (from 1 to 4 expected transitions per branch) is irrelevant for all but the smallest clades. For example, Frumhoff and Reeve (1994: 180) deduce that, for the case in which 4 transitions per branch are expected to have occurred, an inference of an ancestral character state with an error probability of 0.20 or less would require a clade containing a minimum of 2,241 species fixed for the same state. However, the probability of 2,241 species becoming fixed for the same state given 4 expected changes per branch is  $6.2 \times 10^{-675}$ . To put this into perspective, let us assume that the earth's current biota contains on the order of  $5.0 \times 10^7$  species. Even if the history of life were to be reiterated 1,000,000,000 times, the chance of observing even one clade of 2,241 species fixed for the same character state given these conditions is approximately 1 out of  $10^{661}$ .

The only expected numbers of transitions per branch that are likely to produce fixation in a large clade are values close to 0, when the case of homology predominates. (Exceptions can occur in special cases where the character-state transition probabilities are strongly biased in favor of one state; see below.) Figure 2 illustrates the probability of the observed fixation of a character state, as a function of the expected number of transitions per branch, for 2 to 20 species. Figures 3.1 and 3.2 illustrate the maximum probability of erroneously inferring an ancestral state, given that all N descendants are found fixed for one state, evaluated at  $100(1-c)\%$  confidence bounds for expected number of transitions ( $c = 0.05, 0.10, 0.20$ ). These results demonstrate that the phylogenetic inference of ancestral character states is

robust under a Markov chain model with conditions identical to those specified by Frumhoff and Reeve (1994: 180, Figure 2).

Pseudophylogenetic patterns produced by concerted homoplasy

We have shown that the circumstance of multiple transitions per branch, considered by itself, is unlikely to lead to error in ancestral state reconstruction in all but the smallest clades. However, Frumhoff and Reeve (1994) discuss an additional circumstance that, when combined with multiple transitions per branch, will increase the probability of error: that in which conditions strongly favor the non-ancestral state in the descendants. Under the Markov model, this condition is a function of two variables: 1) the probability that the ancestor possessed a particular character state, and 2) an asymmetry in the state-transition probabilities that favors a derived state that was improbable in the ancestor. In Figure 4 we present error curves for the particular case of 8 species in which both quantities are varied (see also Table 1, and Appendix for derivation). Again, when the probability of homoplasious character-state fixation for 8 species is taken into account, a large portion of the calculated error curves (corresponding to probabilities of fixation less than 0.05) are likely to be irrelevant.

It is true, however, that when  $N$  is sufficiently small and when the probabilities strongly favor the non-ancestral state in descendant species, the minimum probability of fixation for the non-ancestral state can exceed 0.05. In these cases, the method employed here is unable to identify a reasonable upper bound on error probabilities for ancestral-state assignments. However, for every hypothetical case of this kind there is a symmetrical case in which the probabilities strongly favor the same state in the descendants as in the ancestor, and in which we would not be



in error. Extrapolating from either case to character evolution in general requires that, whenever a cladogenetic event occurs, character states that are opposite to those in the ancestral species are either universally favored (consistently producing erroroneous inferences) or universally disfavored (consistently producing correct inferences) in the descendants, and that this trend continues only until the next cladogenetic event, so that equilibrium conditions (consistently producing correct inferences) are never achieved.

Frumhoff and Reeve (1994) suggest that the same state might be universally favored in all the descendants of a common ancestor "if closely related species have also experienced a similar series of selective environments and undergone, therefore, a positively correlated sequence of character-state transitions" (p. 173). This is the circumstance of parallelism, widely discussed as a problem for phylogenetic studies (e.g., Hennig, 1966; Eldredge and Cracraft, 1980; Mayr and Ashlock, 1991). All known cases of parallelism have been recognized precisely because the patterns produced by the parallel characters were in some way incongruent with the overriding phylogenetic pattern produced by other characters (e.g., trilobites [Kauffmann, 1933], stalk-eyed flies [Hennig, 1966]). However, Frumhoff and Reeve (1994: 173) suggest a general mechanism of character evolution that is expected to generate parallelisms that exactly mirror phylogenetic patterns. Such undetected parallelisms "might be expected to be fairly common since closely related species tend to share similar niche preferences (Harvey and Pagel, 1991) and responses to selection. Thus, particular character states may be maintained by selection and be widely shared among related species, but be uniquely derived in each of them."

It must be emphasized that, as a substitute for hypothesizing homology for a single character state observed to be shared by a monophyletic group of species, this proposed mechanism for parallelism requires hypothesizing homology for multiple unobserved character states. That is, particular states of two characters rather than one are assumed to be faithfully inherited by each descendant species: the tendency to occupy a similar ecological niche (which we will call  $E_A$ ) and the tendency to respond in a similar way to selection (which we will call  $S_A$ ). In addition, an independent, homoplasious (parallel) origin of the observed character state (e.g., state A in Figure 1) is required in each of the daughter species, as the result of a universal directional selection event acting across all such niches.

Given that a priori information about the rate of character evolution is unavailable, and given that the data consist solely of an observed, uniformly fixed character-state distribution within a clade, we might choose between an hypothesis of homology and an hypothesis of parallelism by comparing their likelihoods, i.e., by comparing the probabilities that they confer on the data. An hypothesis of homology posits that the rate of evolution of the observed character state is low enough that it is improbable that any change has taken place in that character on any of the branches connecting the species within the clade. (So long as the state-transition probabilities are non-integer numbers between 0 and 1, this probability will always be greater than 0 because there is always the chance, however small, that at least one of the observed character states is the product of multiple transitions, e.g.,  $A \rightarrow B \rightarrow A$ .) Thus, an hypothesis of homology confers on the data (consisting of N species fixed for state A) a probability that is close to 1.0.

An hypothesis of parallelism posits that the rate of evolution in the observed character has been rapid enough that it is probable that at least one transition has

occurred on each branch, and that the state-transition probabilities favor the observed state. An hypothesis of parallelism further posits that this bias in transition probabilities applies to all  $N$  species due to their uniform inheritance of two characters,  $E_A$  and  $S_A$ , in which the rate of evolution is low enough that change on any of the branches is improbable. Thus, an hypothesis of parallelism confers on the data (consisting of  $N$  species fixed for state  $A$ ) a joint probability that is the product of, minimally, two quantities: the probability of no change in  $E_A$  and the probability of no change in  $S_A$ . If each of these quantities is large (i.e., close to 1.0), then the likelihood of parallelism may also be large, given the additional requirement of a universal selection event acting across all  $N$  species.

However, for an hypothesis of parallelism to have a higher likelihood than an hypothesis of homology for a given set of data, the joint probability of two independent character-state distributions must exceed the probability of the single character-state distribution associated with homology. For example, if for  $N$  species the probability of fixation for each of the character states  $E_A$  and  $S_A$  is 0.90, and if this is also the probability of fixation for the observed character state (Figure 1), then the likelihood of an hypothesis of parallelism ( $0.90 \times 0.90 = 0.81$ ) is lower than that of an hypothesis of homology (0.90) in accounting for the observed distribution.

In any case, the existence of a correlation between phylogeny and both the tendency to occupy a similar ecological niche and the tendency to respond in a similar way to selection is open to empirical investigation. Also open to investigation is the general claim that entire classes of characters, e.g., "most traits typically studied by evolutionary ecologists" (Frumhoff and Reeve, 1994: 175-176), are especially prone to parallelism. This is at least partly contradicted by the surveys of de Queiroz and Wimberger (1993) and Wenzel (1992), which demonstrate that,

across a wide range of studies, the kinds of behavioral characters utilized by systematists have proven no more homoplasious than morphological ones.

Problems with the model of character evolution employed here

Although we have shown that ancestral character-state optimization performs well under the usual Markov-process assumptions, we caution against generalizing these error probabilities to character evolution as it actually occurs in nature. It is possible that the evolution of some characters may resemble this model (e.g., neutral-region DNA sites); however, the Markov process-based model used here requires assumptions and simplifications that are unlikely to be realistic for many characters, particularly complex behavioral and morphological ones.

Shortcomings of this model include:

- 1) It requires the specification of state-transition probabilities that apply identically and independently to all branches of a phylogeny. As for all probabilities pertaining to biological processes, such probabilities must be obtained empirically (Maynard Smith, 1971: 59). Given a phylogeny, it is possible to retrospectively average the number of transitions between states that occurred in evolution, but it is not obvious how to interpret a probability thus obtained for any particular character. For a character-state transition that occurs more than once on a phylogenetic tree, the empirically obtained state-transition probability may be simply an a posteriori summary of unconnected events. Alternatively, it may be an indicator of an underlying mechanism, e.g., of some innate "quality" of the character, that allows us to predict its behavior in independent evolutionary lineages, even though parallel character-state transitions are not homologous. It seems more likely that character-

state transitions, when they occur, are driven by a multiplicity of processes that vary radically between and within lineages at different points in time.

2) Character states present in outgroup species are not incorporated; however, they also provide important information on rates of evolution within that character, and hence on the probability of error in ancestral-state inference. For instance, the potential error rate for the case of  $N = 2$  species is high because it is not improbable for a rapidly evolving character to be fixed for the same state in two independently evolving species at a given point in time due simply to chance. However, if  $M$  outgroup species are additionally observed to be fixed for the alternate state, and if  $M$  is sufficiently large, then it becomes correspondingly less likely that the character conforms to a rapidly evolving model. Essentially, inferences about the overall rate of evolution will be more accurate if the probability of fixation (Equation A2) includes information from both ingroup and outgroup species.

3) A two-state character provides the worst case for ancestral state reconstruction, because as the number of possible states increases, the probability of fixation for a particular state in  $N$  species decreases even when the transition probabilities favor that state. Given that different species can respond to the same selection event in novel ways, this limitation is clearly unrealistic and inflates the error rate.

4) The polytomous topology (Figure 1) represents the best case for ancestral character state reconstruction, because each observation is conditionally independent and yields maximum information about the ancestor. In the case of more resolved topologies, observations of states in closely related species will be

positively correlated under the Markov model, effectively reducing the sample size of data points bearing on the ancestral state.

## **Conclusion**

Despite these shortcomings, the Markov model has proven useful here in allowing us to further explore the probability of error associated with the reconstruction of ancestral states first proposed by Frumhoff and Reeve (1994). Given the model's assumptions, we have shown that the observation of fixation in a clade for a given character state provides information on the evolutionary rate of change for that character, and that this critical information can be used to place an upper bound on the error associated with ancestral-state inference. Calculation of error using this method demonstrates that, when the clade contains more than 4 species and the state-transition probabilities are not strongly biased in favoring a particular state, ancestral states can be reconstructed with a high degree of reliability.

We hope that additional modifications of the model will lead to increasingly realistic assessments of the error associated with ancestral state reconstruction. In turn, reliable reconstructions of ancestral character states can provide essential information about the sequences of transitions that have occurred in characters that are of particular interest to behaviorists and ecologists, allowing the elucidation of general evolutionary principles and permitting the testing of particular evolutionary hypotheses.

## **Acknowledgments**

We thank A. Brower, J. Carpenter, B. Danforth, M. Engel, D. Maddison, M. McDonald, K. Nixon, K. Reeve, K. Shaw, K. Sime, C. Tauber, M. Tauber, W. Wcislo,

J. Wenzel, and two anonymous reviewers for comments on various drafts of this manuscript. We especially thank C. Foran and J. Brooks for organizing the Neurobiology and Behavior Graduate Seminar group in which these issues were first raised. This paper is dedicated to the memory of George Eickwort, whose commitment to the phylogenetic study of behavior will continue to inspire us.

## **Bibliography**

- Ax, P. 1987. *The Phylogenetic System*. John Wiley, New York.
- Baum, D.A. and A. Larson. 1991. Adaptation reviewed: a phylogenetic methodology for studying character macroevolution. *Systematic Zoology* 40: 1-18.
- Brooks, D. R. and D. A. McLennan. 1991. *Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology*. Chicago: University of Chicago Press.
- Carpenter, J.M. 1989. Testing scenarios: wasp social behavior. *Cladistics* 5: 131-144.
- Casella, G. and R. C. Berger. 1990. *Statistical inference*. Wadsworth, Pacific Grove, CA.
- Chapin, J.P. 1917. The classification of the weaver birds. *Bulletin of the American Museum of Natural History* 37: 243-280.
- Coddington, J. A. 1988. Cladistic tests of adaptational hypotheses. *Cladistics* 4: 3-22.
- Darwin, C. 1859. *On the Origin of Species*. London: John Murray.
- Eldredge, N. and J. Cracraft. 1980. *Phylogenetic Patterns and the Evolutionary Process*. New York: Columbia University Press.
- Farris, J.S. 1970. Methods for computing Wagner trees. *Systematic Zoology* 19: 83-92.
- Farris, J.S. 1983. The logical basis of phylogenetic analysis. Pp. 7-36 in N. I. Platnick and V.A. Funk, eds. *Advances in Cladistics*. Columbia University Press, New York.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401-410.



- Felsenstein, J. 1982. Numerical methods for inferring evolutionary trees. *Quarterly Review of Biology* 57: 379-404.
- Fitch, W. M. 1971. Toward defining the course of evolution: minimal change for a specific tree topology. *Systematic Zoology* 20: 406-416.
- Frumhoff, P. C. and H. K. Reeve. 1994. Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. *Evolution* 48: 172-180.
- Harvey, P. H. and M. D. Pagel. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press.
- Heinroth, O. 1911. Beiträge zur Biologie, namentlich Ethologie und Psychologie der Anatiden. *Verh. Ver. Int. Ornithol. Kongr. (Berlin) 1910*: 589-702.
- Hendy, M.D. and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38: 297-309.
- Hennig, W. 1966. *Phylogenetic Systematics*. Urbana: University of Illinois Press.
- Hillis, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. In *Phylogenetic Analysis of DNA Sequences*, pp. 278-294. New York: Oxford University Press.
- Hillis, D.M., J.P. Huelsenbeck, and C.W. Cunningham. 1994. Application and accuracy of molecular phylogenies. *Science* 264: 671-677.
- Kaufmann, R. 1933. Variationsstatistische Untersuchungen über die "Artabwandlung" und "Artumbildung" an der oberkambrischen Trilobitengattung Olenus Dalm. *Abh. Geol. Pal. Inst. Univ. Greifswald* 10: 1-54.
- Lack, D. 1947. *Darwin's Finches*. Cambridge: Cambridge University Press.

- Lanave, C., G. Preparata, C. Saccone, and G. Serio. 1984. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* 20: 86-93.
- Lorenz, K. 1941. Vergleichende Bewegungsstudien an Anatinen. *Journal of Ornithology* 89: 194-294.
- Maynard Smith, J. 1971. *Mathematical Ideas in Biology*. Cambridge: Cambridge University Press.
- Mayr, E. and P. D. Ashlock. 1991. *Principles of Systematic Zoology*, Second Edition. New York: McGraw -Hill.
- Nelson, G. and N. Platnick. 1981. *Systematics and Biogeography: Cladistics and Vicariance*. New York: Columbia University Press.
- Patterson, C. 1982. Morphological characters and homology. In *Problems in Phylogenetic Reconstruction*, pp. 21-74. K. A. Joysey and A. E. Friday, eds. London: Academic Press.
- de Pinna, M. C. C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7: 367-394.
- de Queiroz, A. and P. H. Wimberger. 1993. The usefulness of behavior for phylogeny estimation: Levels of homoplasy in behavioral and morphological characters. *Evolution* 47: 46-60.
- Ridley, M. 1983. *The Explanation of Organic Diversity: The Comparative Method and Adaptations for Mating*. Oxford University Press.
- Sankoff, D. and P. Rousseau. 1975. Locating the vertices of a Steiner tree in arbitrary space. *Mathematical Programming* 9: 240-246.
- Sankoff, D. and R.J. Cedergren. 1983. Simultaneous comparison of three or more sequences related by a tree. In *Time Warps, String Edits, and*

- Macromolecules: The Theory and Practice of Sequence Comparison, pp. 253-263. D. Sankoff and J.B. Kruskal, eds. Reading, Massachusetts: Addison-Wesley.
- Sober, Elliott. 1988. Reconstructing the Past: Parsimony, Evolution, and Inference. Cambridge, Massachusetts: MIT Press.
- Swofford, D.L. and W.P. Maddison. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87: 199-229.
- Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17:57-86.
- Tinbergen, N. 1951. *The Study of Instinct*. New York: Oxford University Press.
- Wenzel, J. W. 1992. Behavioral homology and phylogeny. *Annual Review of Ecology and Systematics* 23: 361-381.
- Whitman, C.O. 1899. Animal behavior. In *Biological Lectures*, Woods Hole, pp. 285-338. C.O. Whitman, ed. Boston: Ginn and Company.
- Wiley, E. O. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: John Wiley and Sons.

## Appendix

Here we modify the Markov model of Frumhoff and Reeve (1994) to incorporate the information about expected number of transitions per branch that is provided by the discovery of fixation for a particular character state within a clade. The method employed here for incorporating this information is only one of a number of possibilities, and we hope this issue will be pursued further. Like Frumhoff and Reeve (1994), we assume  $N$  species related by a polytomy (i.e., star phylogeny; Figure 1); a single two-state character with states A and B and with state transitions modeled as a Markov process; and independent evolution in each branch leading from the common ancestor to a descendant species.

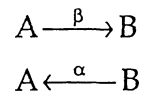
We will assume the usual continuous-time, time-homogeneous Markov process model of evolution (e.g., Tavaré, 1986), such that evolution occurs independently in each branch, with the state-transition probabilities after  $t$  time units given by

$$e^{Qt} = \begin{bmatrix} P_{AA}(t) & P_{AB}(t) \\ P_{BA}(t) & P_{BB}(t) \end{bmatrix}$$

where  $Q$  is the infinitesimal rate matrix given by

$$Q = \begin{bmatrix} -\beta & \beta \\ \alpha & -\alpha \end{bmatrix}$$

and where the relative rates of exchange between state A and state B are



Rather than assume that the process is in equilibrium at the time of divergence from the common ancestor, we define a prior probability that the ancestor was in state B as

$$\Pr(\text{ancestor} = B) = \pi_B$$

The prior probability that the ancestor was in state A is  $1 - \pi_B$ .

We wish to determine the probability that the ancestor possessed state B given that all N descendants possess state A. This probability (as given by Frumhoff and Reeve, 1994) is

$$z(t) = \frac{\pi_B P_{BA}(t)^N}{\pi_B P_{BA}(t)^N + (1 - \pi_B) P_{AA}(t)^N} \quad (\text{Equation A1})$$

where  $z(t)$  is a function of the prior probability  $\pi_B$  and the relative rates  $P_{AB}$  and  $P_{BA}$ , as well as of time  $t$ .

However, this is the error given that we have observed all N descendants fixed for state A, which can be an extremely improbable occurrence when the value of N and/or  $t$  is sufficiently large so that one or more transitions are expected to have occurred on some branches. In such a case, at least some of the descendants share state A due to random homoplasy. In contrast, the fixation of all N descendants for state A will have the highest probability when the ancestor also possessed state A and no changes have occurred on any of the branches. The probability of fixation for state A is

$$F(t) = \pi_B P_{BA}(t)^N + (1 - \pi_B) P_{AA}(t)^N \quad (\text{Equation A2})$$

When confronted with a pattern of fixation in  $N$  species, it is unreasonable to consider that it was generated by values of  $t$  for which  $F(t)$  is low (e.g.,  $F(t) < 0.05$ ). Given that there is a maximum value of  $t$  above which fixation is improbable, the observation of fixation in  $N$  species provides information on the maximum number of transitions that actually have occurred. This upper limit for  $t$  may be identified by setting Equation A2 to a predetermined level (e.g., 0.05) and solving for  $t$ . This approach provides a valid upper confidence bound (Casella and Berger, 1990), which can be calculated for given values of  $N$ ,  $\pi_B$ ,  $P_{BA}$  and  $P_{AB}$ . In the example of Frumhoff and Reeve (1994: 174; their Figure 2),  $\pi_B = 0.5$  and the transition probability matrix at time  $t = 1$  is

$$\begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \quad (\text{Matrix 1})$$

As indicated in our Figure 2, when we consider only those values of  $t$  that are reasonable at given confidence levels ( $F(t) = 0.05, 0.10$ , and  $0.20$ ), as  $N$  increases the expected error due to homoplasy decreases exponentially, and is dramatically lower than that indicated by Frumhoff and Reeve (1994).

In order to model the effects of varying the state transition probabilities ( $P_{AA}$ ,  $P_{AB}$ ,  $P_{BA}$ , and  $P_{BB}$ ) and the prior probability of ancestral state  $\pi_B$ , we consider the case of  $N = 8$  descendant species (Figure 4). For the continuous-time Markov process, the probability of a transition from state B to state A at time  $t$  is given by

$$P_{BA}(t) = \frac{\alpha}{\alpha + \beta} - \frac{\alpha}{\alpha + \beta} e^{-(\alpha + \beta)t}, \text{ while the probability of a transition from state A to state B is given by } P_{AB}(t) = \frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta} e^{-(\alpha + \beta)t} \text{ (Tavaré, 1986). Thus, the model is specified}$$

in terms of 4 parameters:  $t$ ,  $\alpha$ ,  $\beta$ , and  $\pi_B$ . However, in order to avoid confusion between time ( $t$ ) and number of transitions, we will constrain the relationship

between  $\alpha$  and  $\beta$  in such a way that  $\underline{t}$  is expressed in units that are equal to the expected number of transitions per branch.

When the process is in equilibrium,

$$\Pr(\text{ancestor} = B) = \pi_B = \frac{\beta}{\alpha + \beta} \quad \text{and} \quad \Pr(\text{ancestor} = A) = \pi_A = (1 - \pi_B) = \frac{\alpha}{\alpha + \beta}.$$

In this case, the expected number of transitions per branch is given by

$$K = t(\beta\pi_A + \alpha\pi_B) = t \frac{2\alpha\beta}{\alpha + \beta} \quad (\text{Tavaré, 1986}).$$

Thus, by setting  $t \frac{2\alpha\beta}{\alpha + \beta} = 1$ , we can use the constraint  $\beta = \frac{\alpha}{2\alpha - 1}$  to ensure that

$\underline{t} = K$ , i.e., time units are expressed in expected number of transitions, which we will refer to below as  $\underline{t}$ .

If we now define a new parameter

$$\theta_A = 1 - \frac{1}{2\alpha}, \quad 0 \leq \theta_A \leq 1, \quad (\text{Equation A3})$$

the state transition probabilities can be written as

$$P_{BA}(t') = \theta_A - \theta_A e^{-2t'/\theta_A(1-\theta_A)} \quad \text{and} \quad P_{AA}(t') = \theta_A + (1 - \theta_A) e^{-2t'/\theta_A(1-\theta_A)}$$

where  $\theta_A$  describes the tendency of the process to move in the direction of state A, i.e.,  $\theta_A = 0.5$  corresponds to  $P_{AB} = P_{BA}$ ,  $\theta_A > 0.5$  corresponds to  $P_{BA} > P_{AB}$ , and  $\theta_A < 0.5$  corresponds to  $P_{BA} < P_{AB}$ .

In addition to the case of symmetrical exchange between states, two other values of  $\theta_A$  are modeled in Figure 4, one of which favors transitions to state A twice as strongly as the other. For each case, the value of  $\pi_B$  is also varied. Error curves are generated for each case based on Equation 1; for each curve, an upper bound corresponding to a probability of fixation = 0.05 is applied. The results demonstrate that, even when the state transition probabilities favor the state opposite to that possessed by the ancestor, fixation due to random homoplasy is improbable and the error probability associated with ancestral-state inference is low.

However, as the probabilities favoring the opposite state increase, the probability of fixation due to homoplasy also increases, and beyond certain bounds the minimum probability of fixation exceeds 0.05, requiring another method for evaluating relevant error. This result is essentially tautological: when processes strongly favor misleading evidence, we are likely to be misled. It bears noting, however, that for every such case there is a symmetrical case in which the probabilities favor the same state in both the ancestor and descendants, and in which the circumstance of fixation due to random homoplasy rather than homology is highly improbable.



## Figure and Table captions

Figure 1. The polytomous phylogenetic tree ("star phylogeny") modeled here.  $N$  species arise simultaneously from a common ancestor and all are subsequently observed to be fixed for character state A.

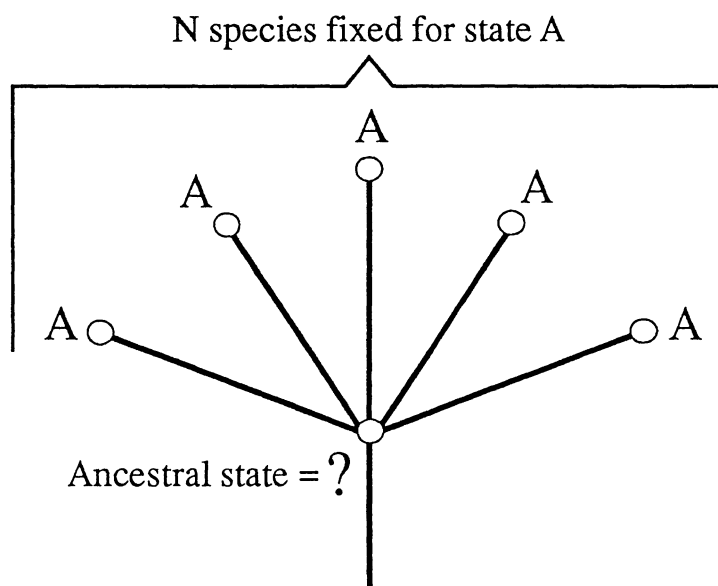
Figure 2. The probability of fixation  $F(t)$  for one state of a 2-state character in  $N$  species, as a function of expected number of transitions per branch (ranging from 0 to 4), calculated using equation A2 and assuming the topology in Figure 1, an ancestral-state probability of 0.5, and the state-transition probabilities in Matrix 1 (Appendix). When the expected number of transitions is low, the probability of fixation is high, with fixation due to homology (i.e., no change) comprising the vast majority of possible outcomes. As the expected number of transitions increases, the probability of fixation decreases, and outcomes due to homoplasy predominate. As the number of species  $N$  increases, the probability of fixation due to homoplasy decreases logarithmically.

Figure 3. The maximum probability of erroneously assigning an ancestral character state (expressed on a log scale in 3.1 and on an arithmetic scale in 3.2), given that all  $N$  species in a monophyletic group of species are observed fixed for one state of a 2-state character. Each curve corresponds to an upper confidence bound (at  $c = 0.20, 0.10, 0.05$ ) on  $\underline{t}$  (time elapsed since divergence), based on the observation of fixation in  $N$  species, i.e., each curve corresponds to the maximum value of  $\underline{t}$  that could have produced the observed fixation with a probability of 0.80, 0.90, and 0.95. When the probability of random fixation in  $N$  species is low, an

observation of fixation indicates homology (i.e., that no transitions have occurred on the branches connecting the ancestor with the descendants). As the number of species increases, the maximum probability of error decreases logarithmically. Assumptions are as in Figure 1; derivation is in the Appendix.

Figure 4. The probability of erroneously assigning an ancestral character state for the case of a clade containing 8 species, given that all are observed to be fixed for state A of a two-state character. Each of the three families of curves corresponds to a different prior probability ( $\pi_B = 0.7, 0.5, 0.3$ ) that the ancestor possessed the opposite state, state B. Each curve within a given family corresponds to a different value of  $\theta_A$  such that transitions favor state A ( $\theta_A = 0.66$ ), transitions favor state B ( $\theta_A = 0.33$ ), or transitions to either state are equiprobable ( $\theta_A = 0.5$ ) (see Appendix). However, only the shaded portions of these curves are relevant at the  $P < 0.05$  confidence level because of the improbability that the number of expected changes (x-axis) in the unshaded region could have generated the observed pattern of fixation. See Appendix for derivation and Table 1 for relevant error values for each curve.

Table 1. The values of  $\underline{t}$  (expected number of transitions per branch) and  $z$  (error in assigning an ancestral state) corresponding to a probability of fixation  $F = 0.05$  for the 9 curves in Figure 4. The portions of the error curves exceeding these values of  $z$  are unlikely to obtain ( $P < 0.05$ ) when a clade of 8 species is found to be fixed for a given character state.



**Figure 1**

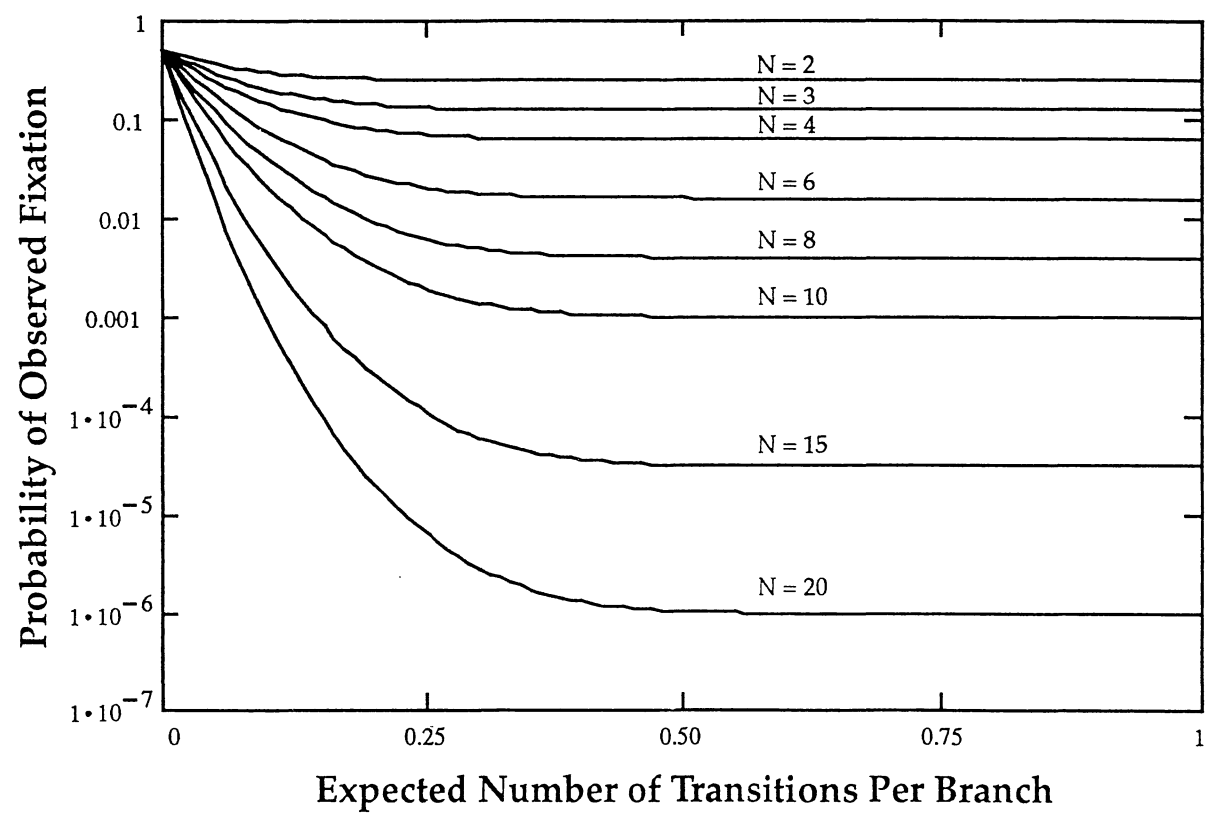


Figure 2

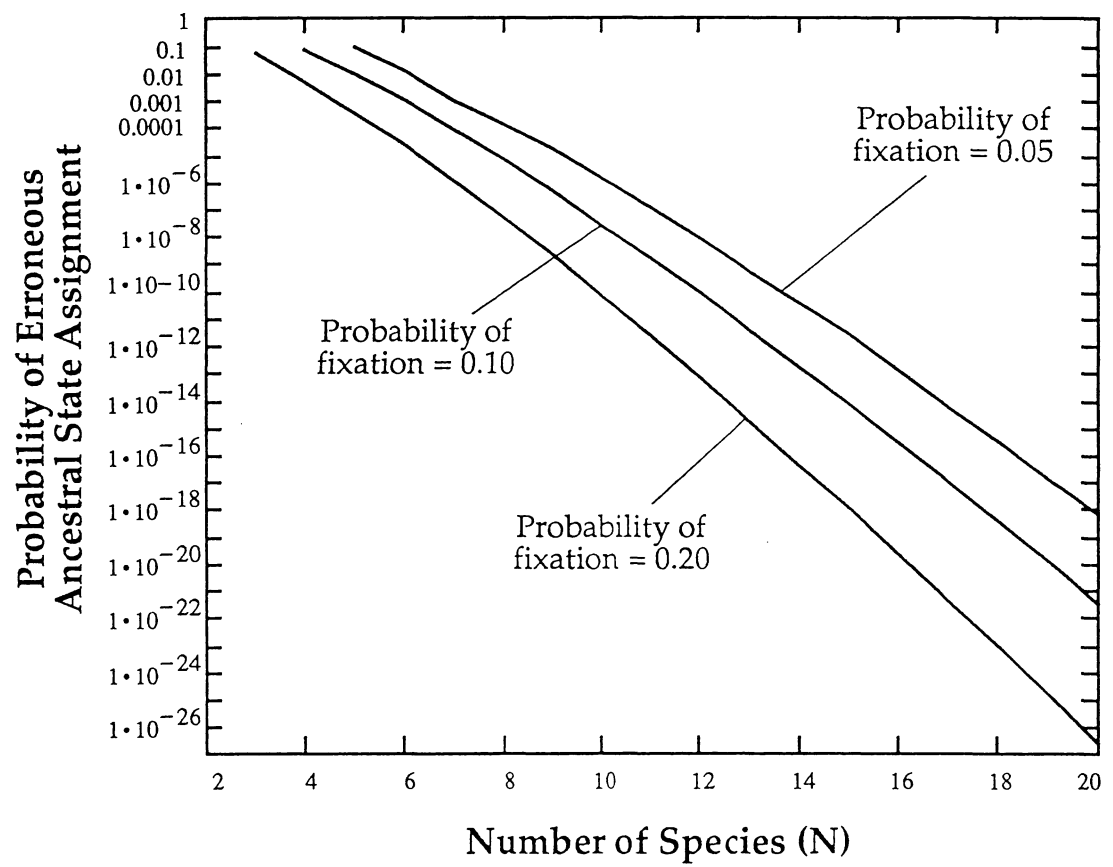


Figure 3.1

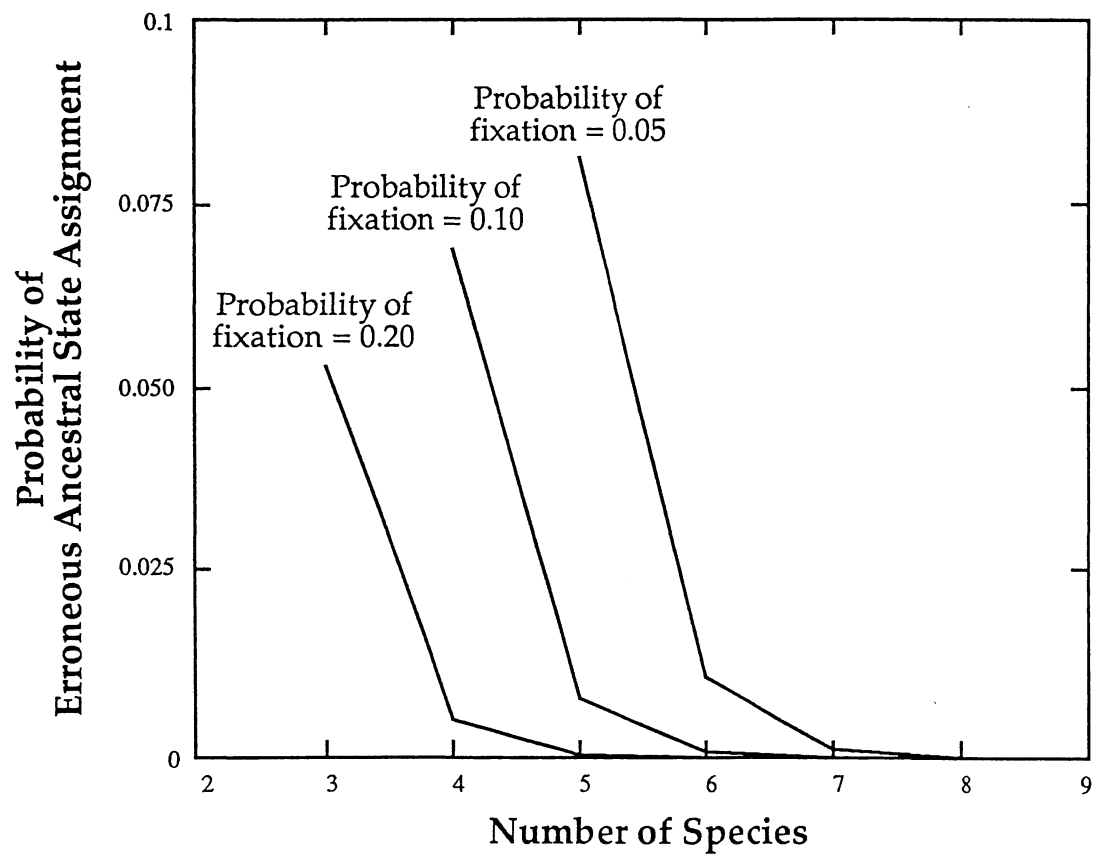


Figure 3.2

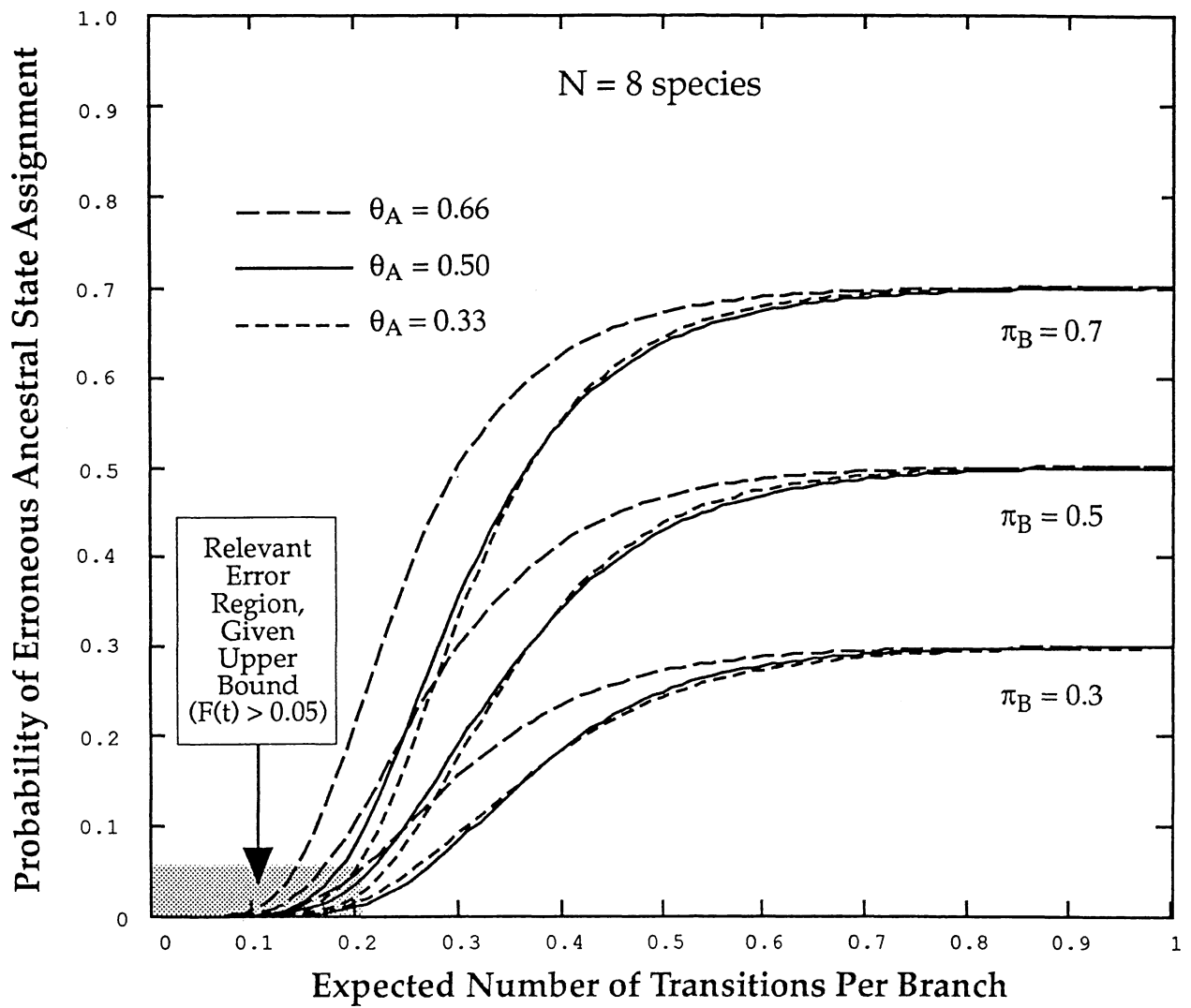


Figure 4

	$\theta_A = 0.66$	$\theta_A = 0.5$	$\theta_A = 0.33$
$\pi_B = 0.3$	$t' = 0.208$ $z = 0.056$	$t' = 0.103$ $z = 2.301 \times 10^{-4}$	$t' = 0.06$ $z = 7.961 \times 10^{-7}$
$\pi_B = 0.5$	$t' = 0.154$ $z = 0.0035$	$t' = 0.087$ $z = 1.571 \times 10^{-4}$	$t' = 0.052$ $z = 4.843 \times 10^{-7}$
$\pi_B = 0.7$	$t' = 0.1$ $z = 0.007$	$t' = 0.064$ $z = 3.623 \times 10^{-5}$	$t' = 0.039$ $z = 1.228 \times 10^{-7}$

Table 1